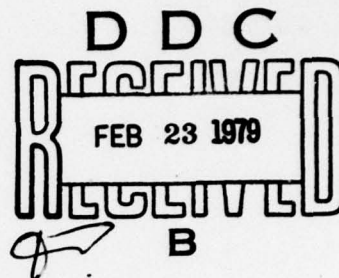② LEVEL ⫪

1

NONSTOCHASTIC TECHNIQUES FOR SELECTING RIDGE

PARAMETER VALUES

Richard F. Gunst and Tsushung A. Hua

DDC
RECEIVED
FEB 23 1979
B

ABSTRACT

Biased regression estimators are increasingly being utilized

as alternatives to least squares parameter estimators in multiple

linear regression when the predictor variables are multicollinear.

One popular biased estimator is the ridge regression estimator.

Ridge estimators are known to have smaller mean squared errors

than least squares for suitably small nonstochastic choices of the

ridge parameter. To date, however, most of the practical applica-

tions of ridge regression employ stochastic techniques to select

the ridge parameter. In this paper we examine three nonstochastic

procedures for choosing ridge parameters and compare their perform-

ance with a stochastic method proposed by Hoerl, Kennard, and

Baldwin (1975).

KEY WORDS

79 02 16 060

NONSTOCHASTIC TECHNIQUES FOR SELECTING RIDGE

PARAMETER VALUES

Richard F. Gunst and Tsushung A. Hua*

1. INTRODUCTION

Hoerl and Kennard (1970a, b) introduced ridge regression estimation as an alternative to least squares estimation of the parameters of a multiple linear regression model when the predictor variables are multicollinear. Write the regression model as

$$\underline{Y} = \beta_o \underline{1} + X\underline{\beta} + \underline{\varepsilon} \qquad (1.1)$$

where $\underline{Y}$ is an $n \times 1$ vector of response variables, $\underline{1}$ is an $n \times 1$ vector of ones, $X = [\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p]$ is an $n \times p$ full column rank matrix of nonstochastic predictor variables that are standardized so that $\underline{X}_j'\underline{1} = 0$ and $\underline{X}_j'\underline{X}_j = 1$ ($j = 1, 2, \ldots, p$), β_o and $\underline{\beta}$ are unknown constants, and $\underline{\varepsilon}$ is an $n \times 1$ vector of random error terms with $\varepsilon_i \sim NID(0, \sigma^2)$. The least squares and (simple) ridge estimators of $\underline{\beta}$ are, respectively,

$$\hat{\underline{\beta}} = (X'X)^{-1} X'\underline{Y} \quad \text{and} \quad \hat{\underline{\beta}}_R = (X'X + kI)^{-1}X'\underline{Y} \qquad (1.2)$$

where $k > 0$.

Theoretical properties, including optimality considerations, of ridge estimators have been derived under the assumption that the

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AFOSR-TR-79-0057		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
NONSTOCHASTIC TECHNIQUES FOR SELECTING RIDGE PARAMETER VALUES	Interim rept.
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Richard F. Gunst and Tsushung A. Hua	AFOSR-75-2871

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Southern Methodist University Department of Statistics Dallas, Texas 75275	61102F 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	January 1978
	13. NUMBER OF PAGES
	23

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ridge Regression
Biased Estimation
Multicollinearity
Least Squares

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Biased regression estimators are increasingly being utilized as alternatives to least squares paramter estimators in multiple linear regression when the predictor variables are multicollinear. One popular biased estimator is the ridge regression estimator. Ridge estimators are known to have smaller mean squared errors than least squares for suitably small nonstochastic choices of the ridge parameter. To date, however, most of the practical applications of ridge regression employ stochastic techniques to select the ridge parameter. In this paper we examine three non-

DD FORM 1473
1 JAN 73

404 198

20. Abstract continued.

stochastic procedures for choosing ridge parameters and compare their performance with a stochastic method proposed by Hoerl, Kennard, and Baldwin (1975).

another

selection of the ridge parameter, k, is nonstochastic. Specifically, if the (total) mean squared error of a regression estimator $\tilde{\underline{\beta}}$ is defined to be

$$mse(\tilde{\underline{\beta}}) = E[(\tilde{\underline{\beta}} - \underline{\beta})'(\tilde{\underline{\beta}} - \underline{\beta})] \, , \tag{1.3}$$

Hoerl and Kennard (1970a) proved that a nonstochastic choice of k that is suitably small would insure that

$$mse(\hat{\underline{\beta}}_R) < mse(\hat{\underline{\beta}}) \, . \tag{1.4}$$

If the latent roots of $X'X$ are denoted $\ell_1 \le \ell_2 \le \ldots \le \ell_p$ and the corresponding latent vectors are $\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_p$, a sufficient condition for (1.4) to hold is that

$$0 < k < \sigma^2/\phi_M^2 \, , \tag{1.5}$$

where $\phi_M^2 = \max\{\phi_j^2; \ j = 1, 2, \ldots, p\}$ and $\phi_j = \underline{V}_j'\underline{\beta}$.

Two problems arise with the practical implementation of ridge regression using (1.5). First, the above properties have only been derived for nonstochastic selection rules for k. Second, the bound in (1.5) is a function of the unknown model parameters. Most of the applications of ridge regression methodology, however, employ stochastic estimators of k (e.g. Hoerl and Kennard 1970b; McDonald and Schwing 1973; and Hocking, Speed, and Lynn 1976). To date, little attention has been given to nonstochastic selection rules.

2. THREE NONSTOCHASTIC SELECTION RULES

One of the motivations for adopting ridge regression over
least squares is that the resulting estimates behave more like they
are generated by an orthogonal X matrix than a multicollinear one.
The rules proposed in this section were investigated because they
enable $(X'X + kI)$ to exhibit properties more closely associated
with the identity matrix than does $X'X$.

If X is an orthogonal matrix (standardized as in Section 1),
the following three properties are true of $X'X$:

(i) $|X'X| = 1$,

(ii) $(X'X)^{-1} = I$ and, therefore, the variance infla-
tion factors (diagonal elements of $(X'X)^{-1}$,
Marquardt (1970)) are all equal to 1 ; and

(iii) the latent roots of $X'X$ are all equal to 1 (i.e.,
$\ell_j = 1$ for all j).

These characteristics all follow from the realization that $X'X = I$
if the columns of X are mutually orthogonal. On the other hand,
if X contains highly multicollinear columns (e.g. Mason, Gunst,
and Webster 1975),

(i) $|X'X| \approx 0$;

(ii) the variance inflation factors are much larger
than 1; and

(iii) one or more of the latent roots of $X'X$ are close
to 0.

In order to compensate for the deleterious effects of multi-collinearities on these properties of X´X, we propose to examine the following three rules for selecting k. The resulting ridge regression estimators, all of which are found using $\hat{\underline{\beta}}_R$ in (1.2), are identified as RR(2), RR(3), and RR(4):

RR(2): Choose k so that $\left|X´X + kI\right| = 1$.

RR(3): Choose k so that the largest variance inflation factor equals 4.

RR(4): Choose k so that the "multicollinearity allowance" (Vinod 1976)

$$m = p - \sum_{j=1}^{p} \ell_j (\ell_j + k)^{-1} \tag{2.1}$$

equals the number of multicollinearities existing among the predictor variables.

The multicollinearity allowance, m in (2.1), can be interpreted (Vinod 1976) as "the assigned deficiency in the rank of X´X." In our investigations if X is an orthogonal matrix, or nearly so, a rank deficiency of zero is assigned to X´X. Solving (2.1) for k when m = 0 yields k = 0; i.e., the ridge estimator in (1.2) reduces to the least squares estimator. On the other hand, if r multicollinearities are identified in X, m is set equal to r and (2.1) is solved iteratively for k.

3. MEAN SQUARED ERROR COMPARISONS

Comparison of stochastic estimators of the ridge parameter have been conducted by simulation due to the complexity of the theoretical distribution of $\hat{\underline{\beta}}_R$ when k in (1.2) is random. Hoerl, Kennard, and Baldwin (1975) estimated k with

$$\hat{k} = p\hat{\sigma}^2/\hat{\underline{\beta}}'\hat{\underline{\beta}} , \qquad (3.1)$$

where $\hat{\underline{\beta}}$ and $\hat{\sigma}^2$ are least squares estimators of the respective parameters. They concluded that $\hat{\underline{\beta}}_R$ using the random estimator of k offered great potential for reducing the mean squared error of $\hat{\underline{\beta}}_R$ over that of $\hat{\underline{\beta}}$. Following up on this study, Hoerl and Kennard (1976) iterated with (1.2) and (3.1) using $\hat{k}$ to obtain $\hat{\underline{\beta}}_R$, then inserting $\hat{\underline{\beta}}_R$ in (3.1) to obtain a new $\hat{k}$, and repeated the process until the estimates $\hat{k}$ (and $\hat{\underline{\beta}}_R$) converged. Their simulation showed that the iterated ridge estimator offered further improvement in mean squared error over least squares and the noniterated ridge estimator. In both of these studies the magnitude of the reduction in mean squared errors of the ridge estimators over least squares was linked to the magnitude of the signal-to-noise ratio, $\rho = \underline{\beta}'\underline{\beta}/\sigma^2$: the larger the value of ρ, the less the reduction in mean squared error of the ridge estimators over least squares. Also, the more multicollinear the data set, the greater the reduction in the mean squared error of the ridge estimators.

Dempster, Schatzoff, and Wermuth (1977) performed an extensive
simulation involving 57 different estimators (including least
squares and several ridge estimators) and 160 model configurations.
Although no replication of the model configurations was reported,
the ridge estimators generally were superior to least squares and
at least competitive with the other biased estimators examined; in
fact, one ridge estimator (RIDGM) was considered the best overall
when comparisons were made with mean squared error, (1.3), as the
criterion.

Gunst and Mason (1977) compared five regression estimators,
including least squares and a ridge estimator using $\hat{k}$ in (3.1).
Their simulation involved 24 model configurations, each of which
was replicated 100 times. Overall, their conclusions parallel the
above ones with two reservations:

> (i)   there are specific model configurations for which
>   the ridge estimator is empirically inferior to
>   least squares, and
> (ii) the ridge estimator was not judged the best overall
>   although it was generally superior to least squares.

One of the reasons pointed out for the latter conclusion was the
high degree of variability associated with the estimator of k --
another motivation for studying nonstochastic selection rules for
k.

We now wish to examine the mean squared errors of ridge regression estimators using choices of k based on the rules RR(2), RR(3), and RR(4). For comparison purposes, we calculate $mse(\hat{\underline{\beta}}_R)$ using (1.3) for the twelve model configurations for the multicollinear X matrix in Gunst and Mason (1977). This 30×10 matrix contains a single strong multicollinearity among the first four columns of predictor variables. The latent roots of $X'X$ for the data are

$\ell_1 = 0.00362 \quad \ell_2 = 0.37145 \quad \ell_3 = 0.49723 \quad \ell_4 = 0.63691 \quad \ell_5 = 0.72105$

$\ell_6 = 1.08692 \quad \ell_7 = 1.22693 \quad \ell_8 = 1.42364 \quad \ell_9 = 1.55308 \quad \ell_{10} = 2.47918.$

The magnitude of the first latent root of $X'X$ compared with the remaining ones indicates that a single strong multicollinearity exists in X. The large elements of $\underline{V}_1$ identify which variables are involved in the multicollinearity:

$$\underline{V}_1' = (0.658, -0.487, -0.451, -0.353, -0.006,$$
$$0.001, -0.003, 0.019, -0.022, -0.022).$$

This latent vector reveals that the first four predictor variables are involved in the multicollinearity. The variance inflation factors (VIF) corroborate the conclusion:

Variable:	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
VIF:	120.1	66.4	57.1	35.9	1.2	1.2	1.3	1.6	1.2	1.6

The twelve model configurations incorporating this X matrix are identified by all combinations of four signal-to-noise ratios, $\rho = \underline{\beta}'\underline{\beta}/\sigma^2$, and three orientations of $\underline{\beta}$ with $\underline{V}_1$, $\phi_1 = \underline{V}_1'\underline{\beta}$. The four signal-to-noise ratios studied are $\rho = 0.04$, 1.0, 100, and $10,000$. The three orientations are determined by letting $\underline{\beta}$ equal $\underline{V}_{10}$ ($\phi_1 = 0.0$), $0.5\{\underline{V}_1 + \underline{V}_2 + \underline{V}_9 + \underline{V}_{10}\}$ ($\phi_1 = 0.5$), and $\underline{V}_1$ ($\phi_1 = 1.0$). These choices of $\underline{\beta}$ enable the coefficient vector to be orthogonal, neither orthogonal nor parallel, and parallel to the vector defining the multicollinearity, respectively.

To eliminate scale differences in the comparisons, we tabulate the scaled mean squared errors, $\text{mse}(\tilde{\underline{\beta}})/\sigma^2$. For the ridge estimators with nonstochastic k the scaled mean squared errors can be written

$$\text{mse}(\hat{\underline{\beta}}_R)/\sigma^2 = \sum_{j=1}^{p} \ell_j(\ell_j + k)^{-2} + k^2\rho \sum_{j=1}^{p} (\ell_j + k)^{-2}\phi_j^2 \quad (3.2)$$

where $\phi_j^2 = (\underline{V}_j'\underline{\beta})^2$ (note that both $\underline{V}_j$ and $\underline{\beta}$ are unit length vectors for these model configurations). Least squares scaled mean squared errors are given by

$$\text{mse}(\hat{\underline{\beta}})/\sigma^2 = \sum_{j=1}^{p} \ell_j^{-1} . \quad (3.3)$$

4. DISCUSSION OF RESULTS

Table 1 displays the theoretical mean squared errors for least squares and the optimal mean squared errors for ridge regression for the model configurations discussed in the previous section. The theoretical value for least squares for all model configurations is given by (3.3); i.e., for the latent roots listed in the previous section

$$\sum_{j=1}^{10} \ell_j^{-1} = 287.63 \ .$$

The optimal mean squared error for the ridge estimators is obtained from the "generalized" ridge estimator

$$\hat{\underline{\beta}}_R = (X'X + K)^{-1}X'\underline{y} \ , \tag{4.1}$$

where $K = \text{diag}(k_1, k_2, \ldots, k_p)$, of which the simple ridge estimator is a special case with $k_1 = k_2 = \ldots = k_p = k$. Using the generalized ridge estimator, the mean squared error (1.3) becomes

$$\text{mse}(\hat{\underline{\beta}}_R) = \sum_{j=1}^{p} \{\sigma^2\ell_j(\ell_j + k_j)^{-2} + k_j^2(\ell_j + k_j)^{-2}\phi_j^2\} \ , \tag{4.2}$$

which is minimized when $k_j = \sigma^2/\phi_j^2$. Inserting the optimal k_j into (4.2) produces the scaled mean squared errors for the generalized ridge estimator that are exhibited in Table 1.

[Insert Table 1]

Simple ridge estimators cannot hope to achieve the optimal mean squared errors shown in Table 1. Nevertheless, these optimal values provide a gauge of the maximum reduction in mean squared error that is possible using ridge regression estimators. Note in particular that if ρ or ϕ_1 is small enormous reductions are possible. If both ρ and ϕ_1 are moderate or large, only small reductions in mean squared error can be obtained; on the contrary, in these situations simple ridge estimators can perform much worse than least squares.

Mean squared errors for the ridge estimators RR(2), RR(3), and RR(4) were obtained by iterating on k until each of the criteria were satisfied to four decimal place accuracy in k. Solutions for k are

RR(2): k = .2100   RR(3): k = .0172   RR(4): k = .0165 .

The theoretical mean squared errors for least squares and the ridge estimators for the twelve model configurations discussed earlier are displayed in Table 2 along with the estimated mean squared errors for ridge regression using (3.1) to estimate k that were obtained in the simulation of Gunst and Mason (1977). The latter ridge estimator is labelled RR(1) in Table 2 and is included for comparison purposes.

[Insert Table 2]

Overall the nonstochastic ridge estimators, especially RR(2), offer as large or substantially larger reductions in mean squared error over least squares as does the stochastic version, RR(1). The notable exceptions to this statement occur for $\rho = 10,000$ and $\phi_1 = 0.5$ or 1.0. Between 6- and 35-fold increases in mean squared error over least squares are observed for the nonstochastic ridge estimators with these two model configurations. The stochastic ridge estimator is comparable in mean squared error to least squares when $\rho = 10,000$ and $\phi_1 = 0.5$ and over three times larger when $\rho = 10,000$ and $\phi_1 = 1.0$.

Conclusions to be drawn from this information include the following:

(i) except for extremely large signal-to-noise ratios, nonstochastic ridge estimators RR(2), RR(3), and RR(4) offer at least as great a reduction in mean squared error over least squares as does the stochastic ridge estimator, RR(1);

(ii) RR(2) is overall as good as any of the biased estimators examined in Gunst and Mason (1977);

(iii) when signal-to-noise ratios are large, nonstochastic ridge estimators can still reduce the mean squared error over least squares if the coefficient vector is orthogonal to latent vectors of $X'X$ that identify multicollinearities; and

(iv) if $\underline{\beta}$ is not orthogonal to latent vectors of $X´X$
that identify multicollinearities and if the
signal-to-noise ratio is large, $mse(\hat{\underline{\beta}}_R)$ can be
orders of magnitude larger than $mse(\hat{\underline{\beta}})$.

5. RIDGE PRELIMINARY TEST ESTIMATORS

Since the only situations in which the nonstochastic ridge
estimators were clearly inferior to least squares in Table 2 occur-
red when ρ was large and $\phi_1 \neq 0$, one might consider the following
estimation procedure. For X matrices with a single strong multi-
collinearity (the procedure is readily generalizable to two or
more multicollinearities), test the hypothesis

$$H_o: \underline{V}_1´\underline{\beta} = 0 \quad vs \quad H_a: \underline{V}_1´\underline{\beta} \neq 0$$

with the statistic

$$F = \ell_1(\underline{V}_1´\hat{\underline{\beta}})^2/MSE ,\qquad (5.1)$$

the usual normal-theory test statistic for H_o. Let F_α denote the
appropriate upper-tail 100α% point of an F distribution with 1 and
n-p-1 degrees of freedom. Then let

$$\hat{\underline{\beta}}_{PT} = \begin{cases} \hat{\underline{\beta}}_R & if \quad F < F_\alpha , \\ \hat{\underline{\beta}} & if \quad F \geq F_\alpha \end{cases} \qquad (5.2)$$

where $\hat{\underline{\beta}}_R$ is a nonstochastic ridge estimator (RR(2), RR(3), or RR(4)). The estimator (5.2) is referred to as a preliminary test estimator since its form depends on the outcome of the test of H_o.

Following Bock, Yancey, and Judge (1973), the preliminary test estimator (5.2) can be written as

$$\hat{\underline{\beta}}_{PT} = I_{[o,F_\alpha)}(F) \cdot \hat{\underline{\beta}}_R + I_{[F_\alpha,\infty)}(F) \cdot \hat{\underline{\beta}} , \qquad (5.3)$$

where $I_{[a,b)}(u) = 1$ if $a \leq u < b$ and equals zero otherwise. A straightforward but tedious derivation reveals that

$$mse(\hat{\underline{\beta}}_{PT})/\sigma^2 = \sum_{j=1}^{p} \ell_j^{-1} + k\ell_1^{-1}(\ell_1+k)^{-1}p_3(\lambda)\{k(\ell_1+k)^{-1} - 2\}$$

$$+ kp_1(\lambda) \sum_{j=2}^{p} \ell_j^{-1}(\ell_j+k)^{-1}\{k(\ell_j+k)^{-1} - 2\}$$

$$+ \rho k\phi_1^2(\ell_1+k)^{-1}\{2p_3(\lambda) - 2p_5(\lambda) + k(\ell_1+k)^{-1}p_5(\lambda)\}$$

$$+ \rho k^2 p_1(\lambda) \sum_{j=2}^{p} (\ell_j+k)^{-2}\phi_j^2 , \qquad (5.4)$$

where $p_r(\lambda) = Pr\{F(r, n-p-1, \lambda) < c/r\}$ and $F(r, n-p-1, \lambda)$ is a noncentral F random variable with r and n-p-1 degrees of freedom and noncentrality parameter $\lambda = \ell_1(\underline{V}_1'\underline{\beta})^2/2\sigma^2$.

The effect of the preliminary test on the mean squared errors of the nonstochastic ridge estimators can be seen in Table 3. To varying degrees both preliminary test procedures (using $\alpha = 0.05$

and 0.01) protect against catastrophic increases in mean squared error over least squares for $\rho = 10,000$ and $\phi_1 = 0.5$ or 1.0, particularly for the latter orientation. Balancing the decrease in mean squared error for these two configurations from those reported in Table 2 is a moderate increase in mean squared error for most of the other configurations. It is interesting to observe, however, that all the nonstochastic ridge estimators using the preliminary test procedure with $\alpha = 0.01$ are competitive with RR(1) except for the configuration with $\rho = 10,000$ and $\phi_1 = 0.5$.

6. CONCLUSION

Nonstochastic ridge selection rules are important to ridge regression methodology because virtually all the theoretical properties of ridge estimators must be derived under the assumption that k is nonrandom. In this study three nonstochastic selection rules for ridge parameters were investigated and compared with one stochastic version. The nonstochastic ridge estimators were shown to compete very favorably with the stochastic one when mean squared error is a criterion, although in some specific model configurations all the ridge estimators are inferior to least squares.

In the course of this investigation several additional observations were made that have a bearing on the conclusions drawn herein. First, Vinod's (1976) use of the minimum ISRM (Index of Stability of Relative Magnitudes) as a criterion for selecting an

appropriate value of the multicollinearity allowance performed
erratically. On some X matrices studied the minimum ISRM indicated
a multicollinearity allowance that appeared to be much larger than
one would select from other considerations (e.g. latent roots and
vectors of X'X), a tendency that was also observed by Wichern and
Churchill (1976). Also, the calculation of ISRM often indicates
several local minima and it is not clear whether the absolute min-
imum or the first local minimum should be utilized. Our preference
is to identify the number of multicollinearities in X through the
use of latent roots and vectors of X'X, variance inflation factors,
and pairwise correlations among predictor variables and set m in
(2.1) equal to the number of multicollinearities so identified.

A second observation made during the course of this study is
that modifications of the nonstochastic selection rules may be
desirable depending on the nature of the multicollinearities. For
example, if two or more strong multicollinearities occur in X,
RR(2) can be less effective in reducing mean squared error than
RR(3) or RR(4). This is because $|X'X|$ can be so close to zero that
a large value of k is required to force $|X'X + kI|$ to equal 1.0.
Large values of k do not reduce the variance portion of $mse(\hat{\underline{\beta}}_R)$ as
substantially as do small ones and may thereby negate the advantage
of using ridge regression.

Finally, modifications of the preliminary test procedure may
be possible. These modifications should be aimed at reducing the

mean squared errors for the ridge estimators at the one or two con-
figurations for which they are inferior to least squares in Table 3
while not causing substantial increases in mean squared error at
the other model configurations.

FOOTNOTE

1. Least Squares and Generalized Ridge Regression Scaled

Mean Squared Errors

Estimators	$\rho = 0.04$	$\rho = 1.0$	$\rho = 100$	$\rho = 10,000$
a. $\phi_1 = 0.0$				
LS	287.63	287.63	287.63	287.63
RR	0.04	0.29	0.40	0.40
b. $\phi_1 = 0.5$				
LS	287.63	287.63	287.63	287.63
RR	0.04	0.81	26.38	252.49
c. $\phi_1 = 1.0$				
LS	287.63	287.63	287.63	287.63
RR	0.04	1.00	73.42	268.82

2. Mean Squared Errors of Least Squares and

Ridge Regression Estimators

Estimator	$\rho = 0.04$	$\rho = 1.0$	$\rho = 100$	$\rho = 10,000$
	a. $\phi_1 = 0.0$			
LS	287.63	287.63	287.63	287.63
RR(1)	24.44	34.93	26.79	208.74
RR(2)	6.51	6.51	7.12	67.50
RR(3)	18.90	18.90	18.90	19.37
RR(4)	19.48	19.48	19.48	19.92
	b. $\phi_1 = 0.5$			
LS	287.63	287.63	287.63	287.63
RR(1)	23.20	18.66	41.03	367.36
RR(2)	6.52	6.79	34.44	2,799.34
RR(3)	18.91	19.07	36.01	1,730.44
RR(4)	19.48	19.64	36.35	1,707.20
	c. $\phi_1 = 1.0$			
LS	287.63	287.63	287.63	287.63
RR(1)	29.81	32.29	100.37	801.66
RR(2)	6.55	7.47	103.15	9,670.48
RR(3)	18.93	19.58	87.15	6,843.79
RR(4)	19.50	20.15	86.79	6,750.64

3. Mean Squared Errors of Ridge Preliminary Test Estimators:

$\alpha = 0.05$ (Top Line) and $\alpha = 0.01$ (Bottom Line)

Estimator	$\rho = 0.04$	$\rho = 1.0$	$\rho = 100$	$\rho = 10,000$
		a. $\phi_1 = 0.0$		
RR(2)	77.79	77.79	78.37	135.72
	26.81	26.82	27.41	87.19
RR(3)	87.84	87.84	87.85	88.30
	38.55	38.55	38.56	39.02
RR(4)	88.28	88.28	88.28	88.69
	39.09	39.09	39.09	39.53
		b. $\phi_1 = 0.5$		
RR(2)	77.80	78.08	107.20	1,330.61
	26.82	27.11	56.82	2,090.71
RR(3)	87.85	88.05	108.25	1,042.61
	38.56	38.75	58.20	1,502.61
RR(4)	88.29	88.48	108.50	1,035.02
	39.10	39.28	58.51	1,488.08
		c. $\phi_1 = 1.0$		
RR(2)	77.83	78.83	179.26	420.71
	26.85	27.86	131.40	1,101.83
RR(3)	87.88	88.66	167.77	394.21
	38.59	39.34	116.67	922.65
RR(4)	88.31	89.09	167.47	393.29
	39.12	39.87	116.28	916.50

REFERENCES

Bock, M.E., Yancey, T.A., and Judge, G.G. (1973), "Statistical
    Consequences of Preliminary Test Estimators in Regression,"
    Journal of the American Statistical Association, 68, 109-16.

Dempster, Arthur P., Schatzoff, Martin, and Wermuth, Nanny (1977),
    "A Simulation Study of Alternatives to Ordinary Least Squares,"
    Journal of the American Statistical Association, 72, 77-91.

Gunst, Richard F. and Mason, Robert L. (1977), "Biased Estimation
    in Regression: An Evaluation Using Mean Squared Error,"
    Journal of the American Statistical Association, 72, 616-28.

Hocking, R.R., Speed, F.M., and Lynn, M.J. (1976), "A Class of
    Biased Estimators in Linear Regression," Technometrics, 18,
    425-37.

Hoerl, Arthur E. and Kennard, Robert W. (1970a), "Ridge Regression:
    Biased Estimation for Nonorthogonal Problems," Technometrics,
    12, 55-67.

_____ (1970b), "Ridge Regression: Applications to Non-
    orthogonal Problems," Technometrics, 12, 69-82.

_____ (1976), "Ridge Regression: Iterative Estimation
    of the Biasing Parameter," Communications in Statistics, A5,
    77-78.

Hoerl, Arthur E., Kennard, Robert W., and Baldwin, Kent F. (1975),
    "Ridge Regression: Some Simulations," Communications in
    Statistics, 4, 105-23.

Marquardt, Donald W. (1970), "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation," Technometrics, 12, 591-612.

Mason, Robert L., Gunst, Richard F., and Webster, John T. (1975), "Regression Analysis and Problems of Multicollinearity," Communications in Statistics, 4, 277-92.

McDonald, Gary C. and Schwing, Richard C. (1973), "Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics, 15, 463-81.

Vinod, H.D. (1976), "Application of New Ridge Regression Methods to a Study of Bell System Scale Economies," Journal of the American Statistical Association, 71, 835-41.

Wichern, Dean A. and Churchill, Gilbert A. (1978), "A Comparison of Ridge Estimators," Technometrics, 20, 301-311.